

BIG DATA E ESTATÍSTICAS PÚBLICAS

BIG DATA AND PUBLIC STATISTICS

BIG DATA Y ESTADÍSTICAS PÚBLICAS

 <https://doi.org/10.56238/arev7n12-313>

Data de submissão: 29/11/2025

Data de publicação: 29/12/2025

Marcus André Alves Zimmermann Vieira

Doutor em População, Território e Estatísticas Públícas

Instituição: Escola Nacional de Ciências Estatísticas (ENCE)

E-mail: marcusazimmermann@gmail.com

Orcid: <https://orcid.org/0000-0002-0869-7952>

Lattes: <http://lattes.cnpq.br/6301720011012851>

Andréa Diniz da Silva

Doutora em População, Território e Estatísticas Públícas

Instituição: Escola Nacional de Ciências Estatísticas (ENCE)

E-mail: andrea.silva@ibge.gov.br

Orcid: <https://orcid.org/0000-0001-9116-0162>

Lattes: <http://lattes.cnpq.br/1436215403793574>

RESUMO

O uso de fontes alternativas de dados como big data, mais do que uma opção com potencial para os institutos nacionais de estatística, está tornando-se uma solução primorosa em um cenário de cortes de financiamento na produção de estatísticas públicas e aumento das taxas de não resposta. O presente artigo tem o objetivo de evidenciar as possibilidades por meio de exemplos positivos em que essa estratégia já tem sido utilizada. Além de abordar alguns conceitos de big data, no presente artigo exploram-se iniciativas da ONU em apoio à propagação desse tipo de iniciativa e expõe-se como essa abordagem tem ocorrido pelo mundo, especialmente na América Latina e Caribe.

Palavras-chave: Big Data. Estatísticas Públícas. ONU.

ABSTRACT

The use of alternative data sources such as big data, more than just a potential option for national statistical bureaus, is becoming an excellent solution in a scenario of funding cuts in the production of public statistics and increasing non-response rates. This article aims to highlight the possibilities through positive examples in which this strategy has already been used. In addition to addressing some concepts of big data, this article explores UN initiatives to support the spread of this type of initiative and explains how this approach has occurred around the world, especially in Latin America and the Caribbean.

Keywords: Big Data. Public Statistics. UN.

RESUMEN

El uso de fuentes alternativas de datos como el big data, más que una opción con potencial para los institutos nacionales de estadística, se está convirtiendo en una solución excelente en un contexto de

recortes de financiación en la producción de estadísticas públicas y aumento de las tasas de falta de respuesta. El presente artículo tiene como objetivo poner de relieve las posibilidades mediante ejemplos positivos en los que ya se ha utilizado esta estrategia. Además de abordar algunos conceptos de big data, en el presente artículo se exploran las iniciativas de la ONU en apoyo de la difusión de este tipo de iniciativas y se expone cómo se ha aplicado este enfoque en todo el mundo, especialmente en América Latina y el Caribe.

Palabras clave: Big Data. Estadísticas Públicas. ONU.

1 INTRODUÇÃO

Conexão. Essa é a palavra que tem definido as relações e as atividades no século XXI. Estamos cada vez mais conectados, desde operações simples, como acender uma luz, antes apertando um interruptor, mas hoje podendo verbalizar isso oralmente para uma assistente virtual, até a montar aulas no ensino público, como tem sido proposto pelo estado de São Paulo, no Brasil (Bocchini, 2024). Apesar das críticas que possam existir quanto até onde vai o papel da tecnologia, é inegável os avanços advindos da mesma. Há ainda as críticas quanto à substituição de trabalhadores por máquinas, o que representaria um desemprego estrutural, contudo essa mudança também pode criar novos empregos, gerando novas oportunidades para quem está em constante atualização.

Com tantas possibilidades de simplificação de atividades, acabamos conectados ininterruptamente e cada vez mais cedo. Dados da TIC Kids Online 2024 mostram que 23% dos entrevistados no Brasil tiveram sua primeira interação com a internet ainda na primeira infância, ou seja, até os seis anos (CETIC.br, 2024). A mesma pesquisa em 2015 sinalizava 11% nesse mesmo indicador (CETIC.br, 2024). Isso reflete um avanço no acesso à internet de uma forma geral e, de fato, um adiantamento no acesso de crianças, que é um assunto controverso e não é objeto de estudo do presente artigo, mas é trazido a título de exemplo para expor um contexto.

Esse uso incessante acaba por gerar rastros. Todos os nossos usos geram uma grande massa de dados, que pode ser utilizada para muitas finalidades. O mundo corporativo já percebeu que esses dados possuem valor e podem orientar a entender melhor sua base de clientes, antecipando preferências, o que é um diferencial de mercado para aumentar a venda de produtos e serviços, além de fidelizar consumidores. Esse conhecimento é extremamente valioso para grandes corporações, como o Grupo Globo, por exemplo, que desenvolveu uma barreira de paywall inteligente. Através de estudos preditivos, a empresa mapeia os temas de interesse dos leitores de seus sites e bloqueia o seu acesso em conteúdos específicos, sendo necessário realizar uma assinatura para acompanhar seu conteúdo digital. Com isso, o leitor interessado acaba por assinar, o que rentabiliza para a empresa.

Contudo, não é apenas o Grupo Globo que tem usado esse tipo de inteligência advinda de uma massa de dados. Segundo Toccato (2021), outras empresas também têm utilizado big data para direcionar suas estratégias de negócio, reduzindo custos ou elevando seu faturamento. A Nike tem orientado a personalização de seus produtos através do monitoramento esportivo dos seus consumidores por meio de dispositivos e tecnologias vestíveis, como relógios e pulseiras inteligentes. A Danone usa dados para melhorar sua logística, aperfeiçoando a roteirização e evitando colocar produtos à venda próximo do fim da validade. A Maplink diagnostica com precisão o trânsito, o que permite contornar congestionamentos. A Netflix consegue recomendar conteúdo analisando

preferências e comportamento do usuário na sua plataforma. A Amazon garante uma maior assertividade no seu relacionamento com o cliente a partir do monitoramento do histórico de buscas e comportamento dos consumidores em seu website. O Grupo Pão de Açúcar consegue personalizar ofertas e preferências de seus consumidores a partir dos dados oriundos de seu programa de fidelidade. A Monsanto aumenta seu rendimento e reduz seus custos a partir de seus dados, otimizando projetos de plantio e fazendo recomendações sobre quais sementes plantar e quais as condições necessárias. Esses são só alguns dos exemplos citados por Toccato (2021), mas diversas empresas estão investindo no uso de big data.

Vale pontuar que não são apenas grandes conglomerados empresariais que conseguem usufruir de big data, já que recentemente está vivendo-se uma democratização do acesso a dados. Com o advento da internet, é possível acessar dados de qualquer lugar do mundo. Atualmente, estamos vivendo uma cultura de transparência que reforça a importância de disponibilizar dados originais de forma a permitir que qualquer um possa replicar experimentos e análises. Essa prática dialoga com a evolução da ciência e permite a execução de um de seus princípios: a reproduzibilidade. Adicionalmente, o desenvolvimento tecnológico e, consequentemente, a popularização de computadores pessoais com alto desempenho permitem que análises sejam feitas com grandes volumes de dados, o que mostra a tecnologia como aliada nesse processo e não um gargalo. Há alguns anos era necessário um grande investimento para manipular grandes volumes de dados, adquirindo data centers com servidores de alto desempenho e grande capacidade de processamento. Atualmente, é possível “alugar” infraestrutura tecnológica apenas pelo tempo de uso, por meio de empresas que oferecem estrutura em nuvem, como Google, Amazon e Microsoft.

Seguindo também esse movimento, a iniciativa pública tem notado o potencial desses dados, o que permite a geração de estatísticas públicas, desonerando as fontes tradicionais em um cenário cada vez mais propenso a cortes de gastos e aumento da não resposta. A Holanda, por exemplo, tem usado big data para ajudar em estatísticas de Turismo (Abdulkadri et al., 2016). Já a Austrália usa para estatísticas agrícolas, a Estônia para estatísticas de viagens internacionais baseadas em dados de localização móvel e nos Países Baixos as redes sociais são fonte de estatísticas oficiais, enquanto o Butão utiliza big data para índice de preços (Peres, 2014). Esse também é o caso do Brasil, que realiza web scraping para coleta de preços de passagens aéreas, que compõem o índice de preços ao consumidor (Silva et al., 2019). Haderen et. al. (2021) cita ainda que Alemanha, Austrália, Áustria, Canadá, China, Colômbia, Espanha, Filipinas, Finlândia, França, Holanda, Índia, Irlanda, Itália, Japão, México, Nova Zelândia, Portugal, Suécia, Suíça, Turquia e Ucrânia já usam big data para produzir indicadores relacionados aos Objetivos do Desenvolvimento Sustentável (ODS).

Vale salientar que dados provenientes de fontes alternativas como big data possuem características diferentes daqueles em registros administrativos ou produzidos por meio de pesquisas. Embora venham sendo cada vez mais utilizados na produção de estatísticas, sua definição ainda não está totalmente consolidada e está em constante evolução. Nesse artigo, além de buscar contribuir para a reflexão sobre conceitos e definições de big data, busca-se mostrar um panorama sobre o uso de big data em estatísticas públicas no mundo, especialmente, na América Latina e Caribe, além de expor os esforços das Nações Unidas para promover o uso de big data para estatísticas públicas.

2 BIG DATA E O JOGO DOS Vs

Big data em tradução livre significa “dados grandes”. Segundo Carvalho (2017, p. 11), dado é “qualquer elemento identificado em sua forma bruta que por si só não conduz a uma compreensão de determinado fato ou situação e pode ser apresentado não apenas na forma de números, mas também de palavras, imagens ou sons”. Para MACFEELY (2019), dados podem ser números, mas também texto, som e imagens. Schroeder (2018) afirma que dados são a menor unidade de análise útil de forma semelhante a um átomo, existindo antes mesmo do processo de análise.

Carvalho (2017) indica que informação seria uma evolução do dado, pois essa já estaria organizada e ordenada, o que permite atribuir significado e entender o contexto. Assim, a informação seria o dado trabalhado com valor agregado e, portanto, que pode ser utilizado para tomar decisões. Informação vem da palavra latina *informare*, que significa dar forma a algo (IBIDEM). Assim, pode-se entender que dado é um insumo presente no estágio inicial da produção da informação. Por fim, o último estágio do dado seria o de conhecimento, que inclui, além de contexto e significado, reflexão e síntese, possibilitando maior qualidade na tomada de decisão, pois agrega um refinamento de informações com uma dose de inteligência, associando também experiências e conceitos para elaborar conclusões (IBIDEM).

Já o termo “grande” nos remete a uma noção de tamanho, em que a dimensão retratada seria maior que o usual. No caso de big data, a literatura aponta que essa noção de tamanho é mais bem associada a volume. Entretanto, como ainda será explorado adiante, o conceito de big data vai além do tamanho, não sendo esse a sua única característica definidora (MACFEELY, 2019). Segundo a Comissão Estatística da ONU (2014, p. 2), big data refere-se a “fontes de dados que podem ser, geralmente, descritas como de grande volume, velocidade e variedade de dados que exigem formas inovadoras e econômicas de processamento para melhor percepção e tomada de decisão”¹. Nesse

¹ Tradução livre de “Big Data are data sources that can be – generally – described as: high volume, velocity and variety of data that demand cost-effective, innovative forms of processing for enhanced insight and decision making.”

ponto, vale apontar os três primeiros Vs: volume, velocidade e variedade. Segundo Pitta e Silva (2015), volume refere-se à quantidade massiva resultante do avanço tecnológico de diferentes formas de geração de dados. Já velocidade traz a ideia da rapidez com que os dados estão sendo gerados em tempo real ou quase isso, o que, por sua vez, intensifica o volume. E a variedade remete à diversidade de dados e formatos, incluindo documentos, imagens, vídeos, mensagens, dentre outros (IBIDEM).

O significado de big data tem evoluído e agregado novas características, afinal seus conceitos e definições ainda estão sendo construídos. Hammer et al (2017) acrescentaram duas características: volatilidade e veracidade. Para o autor, a volatilidade pode dar-se a partir da variação de tecnologia ou do ambiente de negócios em que os dados são produzidos, o que pode levar a análises e resultados inválidos, bem como à fragilidade do uso de big data. Enquanto isso, a veracidade refere-se a dados livres de ruído e viés (IBIDEM). Tam e Halderen (2020) apontam justamente essa característica como algo referido à qualidade e precisão dos dados, pois existem problemas frequentes a isso em big data, como viés de cobertura, autosseleção e erros de medição. Para Pitta e Silva (2015), há necessidade de avaliação da acurácia dos dados, logo essas duas qualidades, na verdade, acabam apontando limitações não exclusivas de big data, mas, ao mesmo tempo, apontam cuidados necessários ao usar qualquer tipo de fonte de dados. Os autores ainda acrescentam que volatilidade está relacionada à instabilidade e até inconsistência dos dados ao longo do tempo, o que pode ocasionar problemas de reproduzibilidade. Por exemplo, uma pesquisa realizada hoje no buscador do Google pode não apresentar os mesmos resultados de uma busca realizada no dia anterior, ainda que utilizando os mesmos termos.

Há ainda mais um “V” a ressaltar-se: valor. Esse termo remete à ideia de custo-benefício, pois essa produção deve gerar uma informação útil com um custo que vale a pena (MACFEELY, 2019). Pereira e Chachapuz (2018) apontam a importância dos dados ao referir-se ao seu potencial para solucionar graves problemas públicos, denotando grande valor científico, social e econômico. As autoras citam ainda o uso de dados para a proteção de direitos fundamentais, como a garantia de segurança pública ante o terrorismo.

Diversos autores vêm acrescentando Vs ao definir big data. Além dos seis Vs já citados, Pranimalar et al. (2017), adicionam mais 11, chegando a 17 Vs, sendo 14 oriundos de uma pesquisa bibliográfica e 3 propostos pelas autoras. Os demais termos encontrados na literatura são: validade, ilustrado como autenticidade dos dados, é descrita pelas autoras como qualidade no tocante à precisão dos dados usados para extrair o resultado na forma de informação. Outra qualidade com V é a visualização, que é conceituada como um processo de representação abstrata. Já a viralidade decorre da velocidade de propagação do dado. Para as autoras, trata-se de uma taxa na qual os dados são transmitidos e difundidos por um usuário e recebidos por diferentes usuários para seu uso. A

viscosidade seria a diferença de tempo entre o evento ocorrido e o evento que está sendo descrito, ou seja, o delay desse processo. A variabilidade decorre da diferenciação dos dados, pois os dados chegam constantemente de diferentes fontes e como eficientemente os mesmos são diferenciados entre dados ruidosos ou dados importantes. Outra característica perde o V na sua tradução: local, do original *venue*, já que os diversos tipos de dados têm como origem várias fontes através de diferentes plataformas. Vocabulário é outra característica, mas envolvendo a terminologia dos dados, devido aos modelos e estruturas de dados. Ainda há a vagueza, que se refere à indistinção de existência em um dado, tendo em vista que a imprecisão diz respeito à realidade das informações que sugerem pouca ou nenhuma reflexão sobre o que cada uma poderia transmitir, ou seja, o significado dos dados encontrados geralmente não é claro, independentemente da quantidade de dados disponível. Assim, fecham-se os Vs da pesquisa bibliográfica de Pranimalar et al. (2017).

As autoras propõem ainda 3 Vs originais: verbosidade, voluntariedade e versatilidade. O primeiro refere-se à redundância da informação disponível em diferentes fontes, enquanto o segundo projeta um comportamento voluntário de big data, pois depende da livre concessão da informação por parte do usuário para ser usada de acordo com o contexto. E, por fim, o último envolve a capacidade de ser flexível o suficiente para ser usado de maneira diversa em contextos diferentes.

Outro V sinalizado por Tam e Halderen (2020) trata da vulnerabilidade que consiste no risco de exposição do conjunto de big data a ameaças à segurança cibernética ou à divulgação de informações individuais, o que feriria a confidencialidade de dados. Contudo, os autores apontam que os Institutos Nacionais de Estatísticas (INEs) têm condições de manter a segurança dos dados com a estrutura já existente e já se utilizam de métodos para garantir a divulgação de resultados sem expor algum dado sensível que identifique algum indivíduo.

Esse jogo dos Vs tem acabado por estimular muitos autores em uma corrida a contribuir com mais um V. Segundo Boeira (2020), já chega a 42 o número de Vs associados com big data. Neste momento, começa-se a contestar a contribuição de tais autores nesse processo de construção do conceito de big data, pois há características semelhantes ou mesmo se inicia uma discussão saudável sobre o que são características de fato de big data e aquilo que é desejável para o melhor aproveitamento do dado, independente do seu tipo ou fonte.

Por fim, ainda vale pontuar uma característica que não começa com um V: complexidade, citada também por Pranimalar et al. (2017). Para Pitta e Silva (2015), a complexidade envolve a integração de várias bases de dados. Isso porque algumas fontes de big data possuem um grande volume de dados, mas poucas variáveis a se explorar. Com isso, é necessário adicionar mais variáveis de outras bases de dados por meio de pareamento para enriquecer as análises. Afinal, estatísticas são mais que médias e

totais, pois contam histórias de fenômenos sociais em que há atores que se relacionam e é papel da estatística interpretar essas relações com base em dados. Assim, partindo daquela ideia de estágios do dado explorada em Carvalho (2017), em que o conhecimento é a etapa final, unir diferentes bases, atribuindo contexto e significado, o que permite geração de reflexão e síntese, apesar de parecer complexo, mostra-se de vital importância, pois desse processo que surgem as conclusões que permitem a adequada tomada de decisão. Pranimalar et al. (2017) também reforçam a importância dessa característica, dadas as correlações das variáveis, ou seja, a relação que pode existir entre elas.

Uma crítica que não pode ser esquecida é quanto a certas características que são propostas pelos Vs, mas não são exclusivas a big data, logo não são suas definidoras, como versatilidade. Somado a isso, há a questão de certos Vs representarem características desejáveis para garantir uma qualidade do dado, mas isso acaba por novamente não definir big data, como valor. Por fim, com tantos Vs, não é de admirar-se que haja uma certa prolixidade e até alguns Vs que se confundem com outros, causando uma sensação de repetição do V, como é o caso de veracidade com validade.

3 O COMITÊ DE ESPECIALISTAS EM BIG DATA DAS NAÇÕES UNIDAS

As Nações Unidas têm exercido um papel importante na promoção do uso de big data para a produção de estatísticas oficiais por meio do Comitê de Especialistas em Big Data e Ciência de Dados (UNCEBD). Esse mantém à disposição dos institutos nacionais de estatística a Plataforma Global, a qual é composta por dados, métodos e especialistas, para apoiar a colaboração internacional na produção de estatísticas oficiais, usando novas fontes de dados e métodos inovadores. Seu valor principal é a confiança, alcançada por meio da colaboração, revisão por pares e aprovação de todo o trabalho realizado na plataforma. A Plataforma Global está fundamentada em quatro pilares: parceiros confiáveis, dados confiáveis, métodos confiáveis e aprendizado confiável. Entre seus recursos, vale destacar a disponibilização de métodos, que compreende uma biblioteca de procedimentos e algoritmos que podem ser explorados por membros de qualquer instituto nacional de estatística (INE), o que reforça o ambiente colaborativo, além de contar com a presença de desenvolvedores, que oferecem acesso a ambientes de criação de código e exploração de dados, fora o suporte aos usuários para desenvolver novos algoritmos e aplicativos utilizando ciência de dados (UNSTATS, 2022). Vale salientar que o contexto para a criação da Plataforma foi a dificuldade de mensuração dos Objetivos de Desenvolvimento do Milênio (ODM) e a busca de alternativas para facilitar o acompanhamento dos Objetivos de Desenvolvimento Sustentável por meio de suas 167 metas e respectivos indicadores, afinal já que as estatísticas tradicionais não atendiam essa mensuração, trabalhar em cima de uma fonte

alternativa poderia ser uma boa saída, sugerindo até mesmo parcerias com a iniciativa privada (D’Alva e Paraná, 2024a; D’Alva e Paraná, 2024b).

Essa iniciativa ainda conta com oito grupos de trabalho, os quais discutem o uso de dados do Sistema de Identificação Automatizado (*Automated Identification System* - AIS em inglês), que compreendem dados de sensores, que podem ser utilizados para medir o transporte de mercadorias, tráfego dentro dos portos, indicadores econômicos de comércio, emissão de gás carbônico e quantidade de pesca, entre outros; o uso de dados de observação da terra, como imagens de satélite, para melhorar as estatísticas oficiais em uma ampla diversidade de tópicos, incluindo agricultura, meio ambiente, atividade empresarial e transporte; o uso de dados de telefonia móvel, que objetiva o desenvolvimento de métodos para a produção de informações, além de verificações de qualidade de indicadores relacionados aos ODS (Objetivos do Desenvolvimento Sustentável) usando dados de telefones móveis; técnicas de preservação de privacidade, que tem por objetivo explorar, desenvolver e compartilhar novos métodos, técnicas e melhores práticas para trabalhar com big data sem comprometer a privacidade; dados de *scanner*, que além desses, incluem o uso de dados de preços de produtos e serviços obtidos por meio de *web scraping*, principalmente para índices de preços; treinamento, competências e desenvolvimento de capacidade, focado na orientação sobre o desenvolvimento e estabelecimento de uma rede global de instituições para treinamento e capacitação para o uso de big data na produção de estatísticas oficiais; o uso de ciência de dados e o georreferenciamento dos indicadores de desenvolvimento sustentável, que foca o trabalho sobre os ODS, a utilização da ciência de dados e a integração de informação geoespacial e estatística; além de aquisição de dados, cujo foco é a negociação global, coprodução e coinvestimento para acesso a fontes de dados de proprietários de dados privados para estatísticas oficiais (UNSTATS, 2022; UNSTATS, 2024).

Para apoiar os institutos nacionais de estatística (INEs) no uso de big data e ciência de dados para a produção de estatísticas oficiais, foram criados quatro hubs regionais e dois hubs globais. Os hubs regionais são sediados por institutos nacionais de estatística no Brasil, na Indonésia, em Ruanda e nos Emirados Árabes, e servem, respectivamente, à América Latina e Caribe, Ásia e Pacífico, África (exceto norte), e Oriente Médio e norte da África. Os hubs globais são sediados pelo instituto de estatística da China e pelo *Basque Centre for Climate Change* – BC3. Os Hubs buscam apoiar os INEs por meio da oferta de capacitação, desenvolvimento de pesquisas e realização de ações que ajude a ampliar a colaboração entre os INEs (UNSTATS, 2024).

O Hub Regional para Big Data no Brasil está sediado na Escola Nacional de Ciências Estatísticas (ENCE) do Instituto Brasileiro de Geografia e Estatística (IBGE) e é fruto de uma parceria com o Departamento de Assuntos Econômicos e Sociais da ONU (UN-DESA), iniciada em abril de

2021. Seu fluxo de trabalho distribui-se em quatro linhas (UNSTATS, 2022). A primeira para fortalecer os laços e promover a cooperação entre produtores de estatísticas oficiais da região; a segunda para capacitar e fomentar o interesse de jovens estatísticos no uso de big data nas estatísticas oficiais; o terceiro para apoiar pesquisas sobre o uso de big data e ciência de dados; e o quarto e último para organização e realização de seminários e conferências.

Nesse ensejo, as Nações Unidas têm mostrado forte apoio no uso de big data por institutos de todo mundo, fomentando uma rede global de colaboração. Para isso, inclui tanto um suporte central como algo mais local. Além de ajudar com treinamento e supervisão, colabora com estrutura para permitir seu uso adequado. Com isso, é possível desde uma atuação junto a um país bastante experiente até um que esteja em um estágio bem inicial neste amadurecimento de uso de big data para estatísticas públicas.

4 BIG DATA E A PRODUÇÃO DE ESTATÍSTICAS AO REDOR DO MUNDO

Diante dos constantes cortes no financiamento da produção de estatísticas públicas, mais do que uma opção, o uso de big data tem sido a solução para a continuidade e até incremento de novas estatísticas produzidas pelos institutos nacionais de estatística (INEs). Como os tipos de big data são diversos, incluindo dados registrados em medidores de consumo, por sensores rodoviários, nas redes sociais, em dispositivos móveis, imagens de satélites, transações comerciais e financeiras, e internet das coisas, as possibilidades são inúmeras, ainda mais que esta lista é apenas ilustrativa e não pretende ser exaustiva.

MacFeely (2019), além de listar os tipos de big data utilizados por organizações nacionais e internacionais, listou tópicos estudados com o uso de big data para a produção de estatísticas oficiais, utilizando informações do *UN Big Data Project Inventory*. Nos projetos registrados, os tipos de big data mais utilizados por organismos nacionais são *web scraping*, dados escaneados de produtos vendidos no comércio varejista e dados de telefonia móvel. Já os organismos internacionais, relataram principalmente o uso de dados de telefonia móvel e de redes sociais (Tabela 1). As temáticas abrangem preços, população, migração, transporte, mobilidade, localização geográfica, espaço, mercado de trabalho, agricultura, uso da terra, turismo, saúde, energia, meio ambiente, criminalidade, corrupção, pobreza, desigualdade social, riscos de desastres, dentre outros (Tabela 2). Nesse estudo, o autor aponta que vários projetos são especulativos ou aspiracionais, nos quais a fonte de big data ainda não foi identificada ou o acesso aos dados (principalmente de telefones celulares) ainda não foi assegurado e, por isso, as estatísticas de preços usando dados da web estão entre os projetos mais frequentes, pois essa abordagem geralmente apresenta maior possibilidade de acesso aos dados.

Tabela 1 - Tipos de big data utilizados por organizações nacionais e internacionais, 2019

Tipos de dados	Nacional	Internacional
<i>Web scraping</i>	22	4
<i>Scanner</i>	20	1
Telefone móvel	14	18
Mídia social	8	23
Imagen de satélite	6	7
Medidor de consumo	5	1
Cartão de crédito	3	1
Sensor de estrada	5	-
Registros de saúde	5	2
Identificação de navio	2	-
Registros criminais	1	2
Outros	20	31
Total	111	90

Fonte - MacFeely (2019)

Tabela 2 – Tópicos estudados por organizações nacionais e internacionais com o uso de big data, 2019

Tópicos de projeto	Nacional	Internacional
Preços	22	4
População/migração	10	4
Transporte/mobilidade	9	11
Geografia/Espaço	8	7
Mercado de trabalho	7	2
Agricultura/uso da terra	6	4
Turismo	5	1
Saúde/doenças	4	7
Energia/ambiente	4	6
Crime/corrupção	2	4
Pobreza/desigualdade	1	9
Redução de riscos de desastres	-	8
Outros	31	24
Total	109	91

Fonte - MacFeely (2019)

Com tanto potencial, as aplicações de big data vêm atingindo os mundos corporativo, acadêmico e governamental. O monitoramento digital, por exemplo, inclui número de cliques, histórico de buscas, tempo de permanência em uma tela e uma infinidade de outras variáveis. Isso permite o mapeamento de tendências de consumo, revelando preferências e gostos do consumidor que alimentam o saber das empresas sobre o seu cliente. MacFeely (2019) aponta que nosso uso diário de tecnologia, desde consultas de pesquisa em buscadores até nossas postagens em mídias sociais, tem deixado pegadas digitais em todos os lugares e esses dados podem ser compartilhados, cruzados e redirecionados de forma nunca realizada, o que potencializa suas possibilidades de uso para a produção de estatísticas. Além disso, o uso de big data já é realidade em estudos multidisciplinares, nas ciências sociais, na área de segurança pública, na saúde, entre outras áreas (Saldanha et. al., 2021; Silva, 2019; Amaral et. al., 2023; Chiavegatto Filho, 2015; Alves et. al.; 2022; Peixoto, 2022)

Tam e Halderen (2020) citam que se utilizam dados de telefones celulares em estudos de movimentação populacional, dados escaneados de produtos para calcular índices de preços, dados de mídias sociais para análise de sentimento, dentre muitos outros, o que evidencia o potencial desse tipo de dado. Halderen et. al. (2021) registram o uso de big data para a produção de indicadores de desenvolvimento sustentável em 15 dos 17 ODS (Quadro 1).

Quadro 1 – Indicadores de ODS calculados a partir de big data, 2021

ODS	Quantidade de indicadores	Tipos de big data	Países que já produziram
1 - Erradicação da pobreza	5	Telefonia móvel; Imagens de satélite.	Filipinas, China e Ucrânia
2 - Fome zero	6	Telefonia móvel; Imagens de satélite; Dados escaneados; Preços online; Mídia social.	Filipinas e China
3 - Boa saúde e bem-estar	9	Telefonia móvel; Mecanismo de pesquisa online.	Filipinas e China
4 - Educação de qualidade	9	Telefonia móvel; Cursos online abertos e massivos.	Filipinas
5 - Igualdade de gênero	4	Telefonia móvel; Imagens de satélite; Mídia social.	Filipinas
6 - Água limpa e saneamento	3	Imagens de satélite.	Austrália, China, Canadá, Holanda, Alemanha, México e Filipinas
7 - Energia acessível e limpa	2	Imagens de satélite; Medidor de consumo.	Filipinas
8 - Emprego digno e crescimento econômico	5	Imagens de satélite; Dados postais (endereço); Telefonia Móvel; Mecanismo de pesquisa online.	Filipinas
9 - Indústria, inovação e infraestrutura	1	Telefonia móvel; GPS; Dados de trânsito (fluxo de carros).	China, Colômbia, Japão e Filipinas
11 - Cidades e comunidades sustentáveis	9	Imagens de satélite; Telefonia móvel; Sensoriamento remoto.	Filipinas, Áustria, China, Colômbia, Irlanda, França, Suécia, Suíça, Canadá, Finlândia, Índia, Itália, México, Portugal e Alemanha
12 - Consumo e produção responsáveis	4	Imagens de satélite.	Filipinas
13 - Combate às alterações climáticas	3	Telefonia móvel; Imagens de satélite.	Filipinas
14 - Vida debaixo d'água	6	Imagens de satélite; Sistema de identificação automatizado.	Austrália, China e Filipinas
15 - Vida sobre a terra	8	Imagens de satélite.	Áustria, China, Finlândia, França, Alemanha, Itália, Nova Zelândia, Espanha, Ucrânia, Filipinas, México, Japão e Turquia

ODS	Quantidade de indicadores	Tipos de big data	Países que já produziram
16 - Paz, justiça e instituições fortes	8	Telefonia móvel; Mídia social.	Filipinas

Fonte – Adaptado de Halderen et. al. (2021)

Com isso, é perceptível que muitos países estão produzindo estatísticas que ajudam a acompanhar o cumprimento das metas estabelecidas pelos ODS. Nesse ensejo, vale pontuar a importância dos dados oriundos de telefonia móvel e imagens de satélite, fora a presença marcante das Filipinas em todos os ODS com indicadores produzidos. Apenas para os ODS 10 e 17 não há exemplos de países, embora para o ODS 10 haja uma sugestão de uso de dados de telefonia móvel (Halderen et. al., 2021). Isso mostra o quanto big data pode ser um grande aliado na produção de estatísticas públicas.

5 BIG DATA E A PRODUÇÃO DE ESTATÍSTICAS OFICIAIS NA AMÉRICA LATINA E CARIBE

Assim, como em outras partes do mundo, o uso de big data também pode ser observado na América Latina e no Caribe. Silva et. al. (2023; 2024) apresentam uma visão geral do uso de big data na América Latina e Caribe com base em consultas anuais realizadas pelo Hub Regional da ONU para Big Data no Brasil aos institutos nacionais de estatística (INEs) da região, nos anos de 2022 e 2023. Segundo os autores, esse movimento que começou no norte global, já está surtindo efeito no Sul, como, por exemplo, Honduras que alia imagens de satélite com registros administrativos para mapear a pobreza e a desnutrição, Colômbia que usa imagens de satélite para calcular indicadores que identificam espaço aberto urbano, e o Brasil que, por meio do Instituto Brasileiro de Geografia e Estatística (IBGE), realiza coleta de preços na internet, utilizando a técnica de *web scraping*. Os autores apontam que Argentina, Brasil, Chile, Colômbia, Costa Rica, Peru e Uruguai sinalizaram que utilizam fontes e big data para a produção de estatísticas oficiais, destacando-se o uso de imagens de satélite e *web scraping*, o que reflete em uma maior produção de informação geográfica e espacial. Enquanto isso, as maiores dificuldades encontradas no uso de big data para a formulação de estatísticas oficiais são a falta de capacidade científica e tecnológica, a insuficiência de recursos humanos qualificados, a estrutura interna de dados não padronizada, a falta de legislação, a má qualidade dos dados, a falta de política de cooperação e a burocratização para acessar big data. D’Alva e Paraná (2024b) ainda alertam sobre os riscos da mercantilização dos dados e, consequentemente, das estatísticas públicas advindas dessas fontes alternativas, fora o risco da concorrência com a iniciativa privada acerca da produção de informação, já que a informação vem cada vez mais aproximando-se de uma mercadoria, o que gera mais interesse por seu valor de mercado do que pelo bem-estar que pode gerar.

Abdulkadri et al. (2016) investigaram a incorporação de big data na produção de estatísticas oficiais no Caribe, destacando as oportunidades que big data oferece para a região, ao mesmo tempo em que sugeriram maneiras de gerenciar os desafios. Os autores buscaram contribuir para a discussão sobre as possibilidades no uso de big data, além de fornecerem uma plataforma sobre a qual uma estratégia de big data caribenha poderia ser construída. Abdulkadri et al. (2016) mencionam o caso do Haiti que usou dados de telefonia móvel para monitorar o movimento de sua população após o terremoto de 2010, o que ajudou na organização de ajuda após a catástrofe. O país ainda usou dados do Twitter e do HealthMap em conjunto com dados do Ministério da Saúde Pública e da População para estudar o surto de cólera após a mesma catástrofe de 2010, demonstrando que indicadores oriundos de fontes alternativas podem antecipar aqueles avindos de fontes tradicionais. Os autores ainda trouxeram o resultado de uma pesquisa da CEPAL enviada para 23 INEs da região, mas com retorno de apenas 10, sendo aproveitadas apenas 7 entrevistas. Todos relataram desconhecer qualquer iniciativa de big data em seu país e apenas um tinha uma estratégia de big data na sua organização.

Enquanto isso, Peres (2014) argumentou que o uso de big data e dados abertos é indispensável para ampliar a quantidade de dados disponíveis a baixo custo e advertiu que a capacidade de formular políticas públicas sustentáveis pode ser prejudicada se essas novas fontes não forem rapidamente exploradas. O autor expos o caso da Colômbia, onde imagens de satélite estão sendo utilizadas para complementar as entrevistas que integram o censo agrícola nacional, além de medir e monitorizar as culturas de coca. Isso sem contar os dados provenientes de pedágios para melhorar os fluxos de tráfego e como insumo para estatísticas de transportes. Assim, por meio desse exemplo, Peres (2014) mostrou o potencial do uso de big data para a produção de estatísticas públicas. Mesmo assim, pudemos perceber que, no Caribe, esse modelo ainda é bastante incipiente. Em contrapartida, a América Latina já avança em um ritmo mais acelerado e promissor.

6 CONSIDERAÇÕES FINAIS

Apesar do jogo dos Vs contribuir para a reflexão sobre a definição de big data, os Vs mais recentes têm agregado pouco na discussão. Isso acontece porque os novos Vs acabam parafraseando outras características ou introduzem atributos esperados para qualquer fonte de dados por reforçar sua qualidade. Apesar de ainda não haver uma definição unânime, o conceito de big data começa a consolidar-se, assim como o uso do mesmo na produção de estatísticas públicas.

Para ajudar nesse cenário, as Nações Unidas têm adotado uma série de medidas, como constituir o Comitê de Especialistas em Big Data e Ciência de Dados (UNCEBD), que promove a Plataforma Global, os Grupos de Trabalhos Temáticos e os hubs regionais e globais. Essas iniciativas estimulam

os países a desenvolverem as melhores práticas no uso de big data para a produção de estatísticas públicas. É notória a experiência de certos países, como as Filipinas nesse movimento, contudo, algumas partes como o Caribe ainda não apresentam experiência no uso de big data. Isso só reforça a necessidade de continuar ampliando medidas para promover esse desenvolvimento a partes ainda não atingidas. Big data é uma fonte alternativa que tem mostrado o seu valor, seja complementando as fontes tradicionais, seja substituindo as mesmas. Logo, é benéfico incluir todos os países nessa revolução nos dados, que tem permitido alcançar estatísticas públicas de qualidade com menor custo e maior agilidade. Isso permite uma gestão pública melhor, que, consequentemente, rende maiores benefícios à população.

AGRADECIMENTOS

O autor principal agradece ao IBGE e a Capes pelas bolsas de estudo concedidas, o que possibilitou o presente trabalho de pesquisa.

REFERÊNCIAS

- ABDULKADRI, A.; EVANS, A.; ASH, T. (2016). An assessment of big data for official statistics in the Caribbean: Challenges and opportunities. *Studies and Perspectives Series – The Caribbean*. N. 48. Disponível em: <https://repositorio.cepal.org/handle/11362/39853>. Acesso em: 15/04/2024.
- ALVES, L. S.; OLIVEIRA, S. B.; SILVA, S. L. F. C. (2022). Aplicabilidades e vantagens do uso do big data no setor de saúde: um estudo bibliométrico. *Anais do 8º EMPRAD (Encontro dos Programas de Pós-Graduação Profissionais em Administração)*. Disponível em: <https://sistema.emprad.org.br/8/anais/arquivos/74.pdf>. Acesso em: 30/04/2024.
- AMARAL, T. B.; VARGAS, D.; PRATES, F. (2023). Segurança pública na era do Big data. FGV Direito Rio. Rio de Janeiro. Disponível em: <https://repositorio.fgv.br/server/api/core/bitstreams/869f434b-4ea1-4b73-ae69-983cde99e084/content>. Acesso em: 30/04/2024.
- BOCCHINI, B. (2024). São Paulo vai usar IA para elaborar aulas digitais da rede pública. Agência Brasil, São Paulo. Disponível em: <https://agenciabrasil.ebc.com.br/educacao/noticia/2024-04/sao-paulo-vai-usar-ia-para-elaborar-aulas-digitais-na-rede-publica>. Acesso em: 17/04/2024.
- BOEIRA, J. P. D. (2020). Conheça os 42 Vs do big data. Época Negócios. Disponível em: <https://epocanegocios.globo.com/colunas/Changemaker/noticia/2020/08/conheca-os-42-vs-do-big-data.html>. Acesso em: 28/03/2024.
- CARVALHO, D. (2017). Aula II de Gestão da Informação. CEFET-RJ, Agosto, 2017.
- CETIC.br. (2024). TIC Kids Online. Indicadores. Disponível em: <https://cetic.br/pt/pesquisa/kids-online/indicadores/>. Acesso em: 20/11/2024.
- CHIAVEGATTO FILHO, A. D. P. (2015). Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiologia e Serviços de Saúde*, Vo. 24(2), pp. 325–332. Disponível em: <https://doi.org/10.5123/S1679-49742015000200015>. Acesso em: 30/04/2024.
- D'ALVA, O. A.; PARANÁ, E. (2024). Estatísticas Públicas, Big Data e Inteligência Artificial: o caso da Plataforma Global da ONU. *Estudos Avançados*. 38. 349-364. 2024a. DOI: <https://doi.org/10.1590/s0103-4014.202438111.018>. Acesso em: 04/11/2025.
- D'ALVA, O. A.; PARANÁ, E. (2024). Official statistics and big data in Latin America: Data enclosures and counter-movements. *Big Data & Society*, 11(1). 2024b. DOI: <https://doi.org/10.1177/20539517241229696>. Acesso em: 05/11/2025.
- HALDEREN, G. V.; BERNAL, I.; SEJERSEN, T.; JANSEN, R.; PLOUG, N.; TRUSZCZYNSKI, M. (2021). Big Data for the SDGs. Country examples in compiling SDG indicators using non-traditional data sources. *Working Paper Series*. ESCAP Statistics Division. SD/WP/12. Disponível em: <https://www.unescap.org/sites/default/d8files/knowledge-products/SD Working Paper no12 Jan2021 Big data for SDG indicators.pdf>. Acesso em: 24/07/2023.
- HAMMER, C. L.; KOSTROCH, D. C.; QUIROS, G.; STA INTERNAL GROUP. (2017). 'Big Data: Potential, Challenges, and Statistical Implications' [online]. IMF Staff Discussion Note, SDN/17/06. Setembro, 2017. Disponível em: <http://www.imf.org/en/Publications/SPROLLs/Staff-Discussion-Notes>. Acesso em: 12/10/2021.

MACFEELY, S. (2019). The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators. *Global Policy*. 10. 121-133. 10.1111/1758-5899.12595. Janeiro, 2019.

PANIMALAR, A.; VARNEKHA, S.; VENESHIA K. (2017). The 17 V's of big data. *International Research Journal of Engineering and Technology (IRJET)*. Vol. 4, Issue 9. Disponível em: <https://www.irjet.net/archives/V4/i9/IRJET-V4I957.pdf>. Acesso em: 27/03/2024.

PEIXOTO, I. G. (2022). Alternativas para o cálculo do indicador ODS 11.7.1: Proporção da área construída das cidades que é espaço público aberto para uso de todos com base em dados abertos. *Dissertação (Mestrado em População, Território e Estatísticas Públicas)*. Escola Nacional de Ciências Estatísticas. Disponível em: https://ence.ibge.gov.br/images/ence/doc/mestrado/dissertacoes/2022/DISSERTAC_A_O_VERSAO_FINAL_ISIS_PEIXOTO_1.pdf. Acesso em: 30/04/2024.

PEREIRA, M. V.; CACHAPUZ, M. C. (2018). Big data e o conflito entre a utilização dos dados e a proteção à intimidade e a vida privada. *RJLB*, Ano 4, nº 1, p. 1067-1085. Disponível em: https://www.cidp.pt/revistas/rjlb/2018/1/2018_01_1067_1085.pdf. Acesso em: 12/03/2021.

PERES, W. (2014). Big data and open data as sustainability tools. A working paper prepared by the Economic Commission for Latin America and the Caribbean. Economic Commission for Latin America and the Caribbean (ECLAC). Disponível em: https://repositorio.cepal.org/bitstream/handle/11362/37158/1/S1420677_en.pdf. Acesso em: 15/04/2024.

PITTA, M.; SILVA, D. B. N. (2015). Seminário de Metodologia IBGE. Big Data em Estatísticas Oficiais.

SALDANHA, R.F.; BARCELLOS, C.; PEDROSO, M. M. (2021). Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde? *Cad Saúde Colet*. Rio de Janeiro. Vol. 29, pp. 51-58.

SCHROEDER, R. (2018). Big data: moldando o conhecimento, moldando a vida cotidiana. *Matrizes* V.12, Nº 2, maio/ago, p. 135-163. São Paulo, Brasil. DOI: <http://dx.doi.org/10.11606/issn.1982-8160.v12i2p135-163>.

SILVA, G. C. (2019). A questão do Big Data nas Ciências Sociais: Panorama inicial a partir da Big Data & Society. *Anais do Seta*. Vol 9 (1), pp. 43-57.

SILVA, A. D.; HYPOLITO, E. B.; OLIVEIRA, F. L. P.; VIEIRA, M. A. A. Z. (2024). Big Data and Statistical Production in Latin America and the Caribbean: Perspectives from National Statistics Offices. *Regional Statistics*. DOI: 10.15196/RS150106. Disponível em: <https://www.ksh.hu/docs/hun/xftp/terstat/online-first/rs150106.pdf>. Acesso em: 05/03/2025.

SILVA, A. D.; OLIVEIRA, B. M. M.; PEIXOTO, I. G.; SOUZA, L. B. S. (2023). Overview of the Use of Big Data for Official Statistics in Latin America and the Caribbean. *Statistical Journal of the IAOS*, vol. 39, n. 1, pp. 171-177. Disponível em: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji220092>. Acesso em: 16/04/2024.

SILVA, L. T.; OLIVEIRA, I. L.; DANTAS, T. M.; MIRANDA, V. G. (2019). Studies of new data sources and techniques to improve CPI compilation in Brazil. 16th meeting of the Ottawa Group. Disponível em: https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/study_of_new_data_sources_snipc_lincoln_da_silva.pdf. Acesso em: 15/10/2024.

TAM, S. M.; HALDEREN, G. V. (2020). The Five V's, Seven Virtues and Ten Rules of Big Data Engagement for Official Statistics. *Statistical Journal of the IAOS*, vol. 36, n. 2, pp. 423-433. Disponível em: <https://content.iospress.com/download/statistical-journal-of-the-iaos/sji190595?id=statistical-journal-of-the-iaos%2Fsji190595>. Acesso em: 15/04/2024.

TOCATO. (2021). Da Nike à NASA: conheça 10 exemplos de empresas que usam o Big Data. Disponível em: <https://toccato.com.br/blog/da-nike-a-nasa-conheca-10-exemplos-de-empresas-que-usam-o-big-data/#Nike>. Acesso em: 22/11/2024.

UNSD. (2014). Fundamental Principles of Official Statistics. Disponível em: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>. Acesso em: 24/11/2021.

UNSTATS. (2022). Regional Hubs. Disponível em: <https://unstats.un.org/bigdata/regionals-hubs.cshtml>. Acesso em: 09/03/2022.

UNSTATS. (2022). Task Teams. Communities of global partners working together to explore, develop, and share new methods, techniques and best practices for working with big data. Disponível em: <https://unstats.un.org/bigdata/task-teams/index.cshtml>. Acesso em: 09/03/2022.

UNSTATS. (2022). UN Global Platform. Data For the World. Disponível em: <https://unstats.un.org/bigdata/un-global-platform.cshtml>. Acesso em: 09/03/2022.

UNSTATS. (2024). Report of the Committee of Experts on Big Data and Data Science for Official Statistics. New York. Fifty-fifth session. Item 3 (d). Disponível em: https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/2024-6-BigData-E.pdf. Acesso em 30/04/2024.

UNSTATS. (2024). UN Hubs for Big Data and Data Science. <https://unstats.un.org/bigdata/hubs/>. Acesso em: 30/04/2024.